RESEARCH ARTICLE

Machine learning to support social media empowered patients in cancer care and cancer treatment decisions

Daswin De Silva[1]*, **Weranja Ranasinghe**[1,2], **Tharindu Bandaragoda**[1], **Achini Adikari**[1], **Nishan Mills**[1], **Lahiru Iddamalgoda**[1], **Damminda Alahakoon**[1], **Nathan Lawrentschuk**[2], **Raj Persad**[3], **Evgeny Osipov**[4], **Richard Gray**[5], **Damien Bolton**[2]

1 Research Centre for Data Analytics and Cognition, La Trobe University, Victoria, Australia, **2** Austin Hospital, Heidelberg, Victoria, Australia, **3** North Bristol, NHS Trust, Bristol, United Kingdom, **4** Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Luleå, Sweden, **5** School of Nursing and Midwifery, La Trobe University, Victoria, Australia

* d.desilva@latrobe.edu.au

Abstract

Background

A primary variant of social media, online support groups (OSG) extend beyond the standard definition to incorporate a dimension of advice, support and guidance for patients. OSG are complementary, yet significant adjunct to patient journeys. Machine learning and natural language processing techniques can be applied to these large volumes of unstructured text discussions accumulated in OSG for intelligent extraction of patient-reported demographics, behaviours, decisions, treatment, side effects and expressions of emotions. New insights from the fusion and synthesis of such diverse patient-reported information, as expressed throughout the patient journey from diagnosis to treatment and recovery, can contribute towards informed decision-making on personalized healthcare delivery and the development of healthcare policy guidelines.

Methods and findings

We have designed and developed an artificial intelligence based analytics framework using machine learning and natural language processing techniques for intelligent analysis and automated aggregation of patient information and interaction trajectories in online support groups. Alongside the social interactions aspect, patient behaviours, decisions, demographics, clinical factors, emotions, as subsequently expressed over time, are extracted and analysed. More specifically, we utilised this platform to investigate the impact of online social influences on the intimate decision scenario of selecting a treatment type, recovery after treatment, side effects and emotions expressed over time, using prostate cancer as a model. Results manifest the three major decision-making behaviours among patients, Paternalistic group, Autonomous group and Shared group. Furthermore, each group demonstrated diverse behaviours in post-decision discussions on clinical outcomes, advice and expressions of emotion during the twelve months following treatment. Over time, the

transition of patients from information and emotional support seeking behaviours to providers of information and emotional support to other patients was also observed.

Conclusions

Findings from this study are a rigorous indication of the expectations of social media empowered patients, their potential for individualised decision-making, clinical and emotional needs. The increasing popularity of OSG further confirms that it is timely for clinicians to consider patient voices as expressed in OSG. We have successfully demonstrated that the proposed platform can be utilised to investigate, analyse and derive actionable insights from patient-reported information on prostate cancer, in support of patient focused healthcare delivery. The platform can be extended and applied just as effectively to any other medical condition.

Introduction

Online support groups (OSG) are an increasingly indispensable patient-centred resource for all medical conditions and illnesses as research shows that more than 80% of Internet users seek information related to medical or personal problems via online resources and social media platforms [1–3]. The institutional void of a resilient network of support for individuals (patients and partners) in relatable circumstances is the primary reason for this prevalence [4]. Widespread technology availability, literacy, accessibility and opportunity for archival search are some of the secondary reasons [5]. OSG are anonymous comfortable virtual spaces for patients, carers and information seekers to share experiences, seek advice, express emotions and provide emotional support [6–9].

OSG discussions are organized as discussion threads, where each thread starts with a question, comment or an experience about the corresponding patient's health concerns. Other patients on the OSG respond to these concerns, thereby creating discussion threads.

Fig 1 presents an anonymised sample of five OSG posts by a patient, from diagnosis of cancer to four months post-surgery. It demonstrates the wealth of implicit information contained within OSG posts. Patients begin by mentioning demographic and clinical information, followed by their decision-making process, relevant decision factors and emotions, in order to seek validation from other patients [10,11]. The timeline of clinical and emotion information is implicit in the time-stamp of the post and often explicitly mentioned in the post content. However, this entire body of information is encapsulated within large volumes of unstructured text data [12] which lacks a domain-specific structure required for investigation or intervention and support by primary care providers. Advances of machine learning [13–17], deep learning [18–21] and natural language processing [22–26] present an ambitious opportunity for enabling this transition by capitalising on the prevalence of OSG and their respective networks of support.

In this paper, we present the Patient Reported Information Multidimensional Exploration (PRIME) framework for automated investigation of patient behaviours, clinical factors and patient emotions, across the temporalities of diagnosis, treatment and recovery. More specifically, we focus on the automated multi-granular extraction, analysis, classification and aggregation of decision-making behaviours, decision factors, temporality of patient interactions, temporality of clinical information and side effects, and trajectory of positive and negative
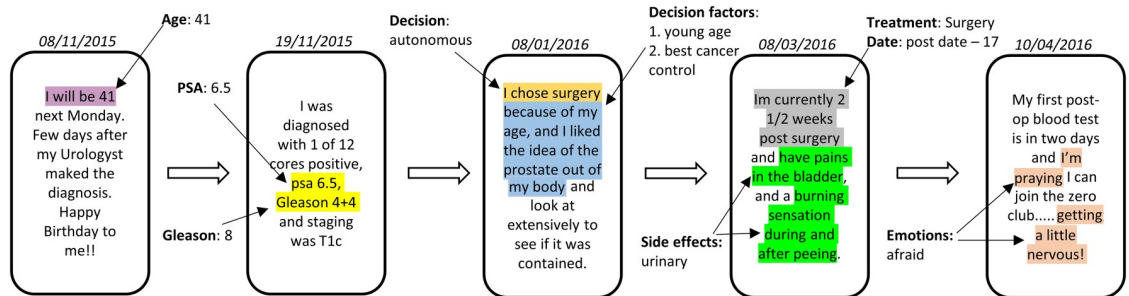
Fig 1. An anonymised sample (parts are omitted and rephrased to preserve privacy) of five posts by a prostate cancer patient. The highlighted excerpts are demographic, clinical, emotion expressions and decision make process related information, stated in the form of free-text.

https://doi.org/10.1371/journal.pone.0205855.g001

emotions, in the context of decision groups, demographics and treatment type. The initial development of PRIME comprised of an ensemble of machine learning (ML) algorithms and natural language processing (NLP) techniques exclusively focused on addressing the nature, content and variety of OSG discussions [27–30]. The NLP techniques map everyday language on to ontology-driven vocabularies and thereby introduce clinical context into informal discussions. The ML algorithms distinguish between diverse patient behaviours and associate these with clinical contexts and patient demographics.

We applied PRIME on OSG for prostate cancer patients. PRIME can be seamlessly applied to any OSG focused on a different medical condition. Our selection of prostate cancer (PCa) was motivated by several factors. They are the complexity of selecting a treatment type for PCa (taking into account cancer maturity and likelihood) [31], PCa has the highest five-year relative survival rate[32], PCa is among the highest reported type of cancer[33], one of the least supported in terms of patient-centred care[34] and most in need of patient education[35].

Related work

Numerous research endeavours have been reported in recent literature for determining patient factors from free text discussions in OSG. A majority of these are qualitative approaches, based on manual categorisation of OSG posts by domain experts. The categorisations include (i) the type of support sought/provided such as emotional/informational/medical/networking [36–38], (ii) the type of emotions expressed [39–41], and (iii) other illness specific topics discussed [38,42]. A key limitation of manual categorisation is that volume is limited to several hundred discussions.

Automated intelligent text analysis methods have been proposed for the analysis of large volumes of discussions. Such methods include unsupervised approaches such as topic capturing [43] and text clustering [44,45] to understand the topics discussed in OSG posts, as well as supervised techniques to capture different categories of OSG posts based on manually coded training datasets [46]. Standard linguistic ontologies [47] to measure emotional and psychological aspects of the OSG posts [48,49] as well as deep learning based classification methods to characterise the mental state of the author based on expression of language [50], have also been reported.

However, all related work is limited to a single aspect of online support, thereby lack the capacity to investigate, analyse and derive actionable insights, over time, from diverse patient-reported information.

Methods

Data collection

OSG data related to PCa was collected from ten high volume active OSG focused on PCa discussions. An active OSG is defined as having at least 100 new conversations per week. From these active OSG, conversations were automatically filtered using the specific topic *'prostate cancer'*. The collected dataset contains 609,960 conversations from 22,233 patients, comprising a text corpus of 93,606,581 word tokens.

Inclusion criteria

Since our interest is in patient decision-making, across different PCa treatment modalities [51], we have set our inclusions criteria as patients who have self-disclosed their chosen PCa treatment and discussed the decision-making process that led to the selection. Note, that PRIME was utilised to automatically extract this information from the collected OSG discussions. A total of 6,457 patients (29%) met these inclusion criteria and thus, selected for this study. Table 1 presents the distribution of patients who met the inclusion criteria across the ten OSG.

Ethical considerations

We have obtained ethics approval for this research from the La Trobe University Human Ethics Committee. All patient-reported data used in this study are non-identifying and publicly available from the corresponding OSG. The OSG does not provide access to identifying information of patients, and we have not processed any identifying information using PRIME. We have only published aggregates of the analysed data, which cannot be reverse engineered using any means for any form of re-identification.

Patient Reported Information Multidimensional Exploration (PRIME)

The PRIME framework functions in seven stages S1-S7 as depicted in Fig 2. Stages S1-S3 are based on our previous work [27–29]. All stages are delineated in the following subsections.

Stages S1-S3

An OSG comprises a large number of discussions where patients contribute their decisions, experiences and opinions at different stages of their patient journey from diagnosis to post-

Table 1. The patient distribution (inclusion criteria met) across the ten selected OSG.

| Online support groups (OSG)	URL	n (% in total)
Healingwell	www.healingwell.com/community	2520 (39.0)
Cancerforums	www.cancerforums.net	873 (13.5)
Cancer Survivors Network	csn.cancer.org/forum	810 (12.5)
Healthboards	www.healthboards.com/boards	429 (6.6)
Prostatecancerinfolink	prostatecancerinfolink.ning.com/forum	396 (6.1)
Cancercompass	www.cancercompass.com	356 (5.5)
Prostatecanceruk	community.prostatecanceruk.org	308 (4.8)
Patientinfo	patient.info/forums	299 (4.6)
Ustoo	www.inspire.com/groups/us-too-prostate-cancer	236 (3.7)
Macmillanuk	community.macmillan.org.uk	230 (3.6)

https://doi.org/10.1371/journal.pone.0205855.t001
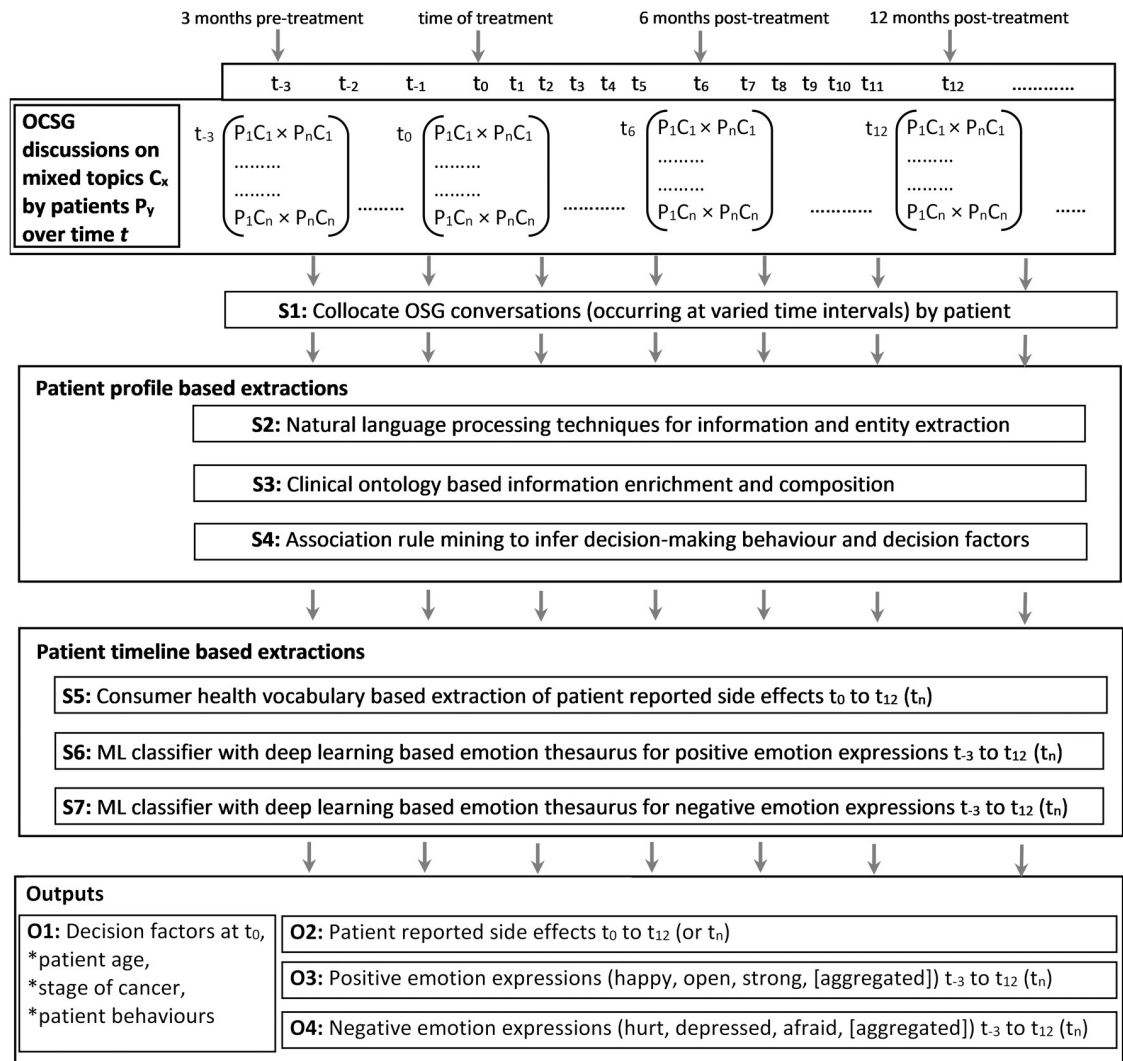
Fig 2. Structural and functional elements of the PRIME framework.

https://doi.org/10.1371/journal.pone.0205855.g002

treatment. The naturally occurring order of discussions provides a multitude of granular and aggregate information on patient behaviours, side effects and emotion expressions over time. However, posts by a single patient are scattered over multiple discussions. Therefore, *S1*, collocates conversations by a single patient, chronologically ordered based on timestamp. In *S2*, NLP based information retrieval techniques [24] are used to process the text corpus and subsequently, machine learning algorithms for classification are utilised to extract demographic information mentioned in free text [28]. Next, *S3* enriches this multidimensional information model with prostate cancer specific clinical information, which are important to categorise patients based on the stage of cancer. In relations to prostate cancer, Gleason and PSA information are key determinants that are extracted. In *S3*, association rules and extracts from clinical ontologies [52,53] are utilised to capture multiple narrative styles for Gleason and PSA mentions(e.g., 'GS3+3', 'Gleason 7'). Subsequently, a classifier based on regular expressions was developed to capture the numerical details of Gleason and PSA scores. Further deliberations can be found in [27–29].

Stage S4

Each patient's decision making behaviour was inferred based on three well-established decision behaviour groups [54,55] (i) Paternalistic: those who strictly adhere to clinician recommendations, (ii) Autonomous: those who are solely driven by personal preference, and (iii) Shared: a mixed group whose decisions are based on both clinician recommendations and personal preferences. We hypothesised this information is encapsulated in the OSG posts which contain mentions of treatment options for prostate cancer.

A set of template patterns was engineered to capture sentences that describe that either individual has taken the decision (Autonomous) or the treatment option was recommended by a clinician (Paternalistic). The template patterns are as follows:

- Autonomous template: <I/We> <words>* <DECIDE> <words>* <TREATMENT>

- Paternalistic template: <DOCTOR> <words>* <RECOMMEND> <words>* <TREATMENT>

Note that <words>* denotes zero or multiple words in-between, and uppercase terms are template terms which consider a set of synonym terms (word/phrase). Table 2 shows a selected sample of terms for each template term.

Multiple decision factors, both clinical/non-clinical affect the treatment decision-making process. As shown in Fig 1, patients often mention decision factors alongside the mentions of the treatment decision. These decision factors were captured using a thesaurus of consumer health terms. Initially, a list of common decision-making factors related to prostate cancer was created based on existing literature [42,56–59] and further validated by clinicians. This list includes medical concerns such as side effects, doctor skills, and best cancer control as well as socio-demographic reasons such as age, fast recovery and financial concerns. Association rule mining was conducted on the corpus to determine decision factors for each patient from this list.

Stage S5

Stage *S5* onwards, PRIME framework incorporates the time dimension of OSG discussions and patient interactions. A patient event timeline is automatically generated for each individual based on the self-disclosed side effects captured in *S5* and positive/negative emotions captured in *S6-S7*. Each patient timeline is time-normalised by considering the treatment month captured in *S4* as t_0. The events (side effects and emotions) are aggregated monthly based on the reported timestamp, and the timeline is generated from three months pre-treatment (t_{-3}) to 12 months post-treatment (t_{12}) based on the available information. *S5* captures the self-disclosure of side effects and grouped into four key categories: *urinary*, *sexual*, *bowel* and *other* which represent side effects of prostate cancer treatments. Note that, *other* represents the miscellaneous side effects such as *hernia*, *clots* etc. A thesaurus of relevant terms (words/phrases)

Table 2. Sample terms for template decision making terms.

Template term	Candidate sample terms
Decide	decided, chosen, wind up going, made the call, settled, opted, went for, took the option, end up
Recommend	recommend, recommended, prescribe, prescribed, advised, advise, endorse, endorsed, advocate
Doctor	doctor, doc, surgeon, urologist, uro, specialist, consultant, radiologist, oncologist, radiotherapist
Treatment	Surgery: surgery, davinci, da vinci, robotic, prostatectomy, ralp, rrp, lrp, rpp, key hole, open op
	Radiation: radiation, imrt, brachytherapy, radiotherapy, seed therapy, brachy, seed implant, ebrt
	Surveillance: surveillance, AS, watch and wait,

https://doi.org/10.1371/journal.pone.0205855.t002

was used to capture any mentioned of an occurrence of side effects and map such mentions to the timeline based on the associated timestamp. Even though the clinical terms for side effects are well defined and recorded in clinical ontologies [52], individuals often describe side effects using everyday language (e.g., urinary incontinence described as leakage, leak, drip), which are not found in clinical ontologies [60]. Therefore, a sample of OSG posts was examined by a team of clinical experts, and consumer health terms related to each side effect category were captured and included in the thesaurus.

Stages S6-S7

As established in the clinical literature, OSG are an accommodative environment for patients to freely express emotions [61,62]. Expressions of emotion reflect Quality of Life (QoL) measures such as living with the condition, the impact of treatment preferences and side effects. In S6, a machine learning technique incorporating a domain-specific vocabulary of positive emotion expressions determines explicit and implicit instances of positive emotions, emotion categories and associated strength of emotion and in S7, this was extended to negative emotion expressions.

Many psychological emotional models have been proposed in the research literature to represent human emotions. These range from the two-dimensional valence-arousal model [63] to multi-dimensional models such as emotion wheel [64]. While such models serve as the theoretical basis for emotion representation, computational implementations must capture expressions of emotion from textual discourse. For example, sentiment analysis techniques are the computational implementation of the valence-arousal model [65], which provide a signed real-value as the sentiment score, where the sign (positive/negative) represents the valence and the absolute value of score represents arousal. Although sentiment analysis techniques are relatively mature and commonly used for capturing emotions, the two dimensional model is coarse-grained for representing complex emotional states of OSG users. Therefore, we developed a a new machine learning technique based on the Emotion Wheel [64,66] to capture a multi-dimensional representation of emotions.

Emotion Wheel has eight primary emotions (joy, trust, surprise, sadness, disgust, anger, anticipation and fear) and further eight secondary emotions which are derived using combinations of primary emotions (e.g., love: joy+ trust). These 16 emotions (primary and secondary) specified in the Emotion Wheel were incorporated as the emotional dimensions in the proposed computational model. The emotional intensity of each emotion is determined based on the proportion of relevant emotional terms present in each OSG post, resulting in a 16-dimensional real-valued emotion vector for each OSG post.

Fig 3 presents the implemented technique for emotion extraction. The relevant terms for each emotion are obtained using a two-step process. First, a seed emotion term thesaurus is constructed for each emotion based on a list of feeling words used for mental status exams [67], which contains emotion terms for each of the 16 emotions.

Expanding a seed list of lexicons is a tedious activity, which is often achieved using crowd-sourcing techniques such as Amazon Mechanical Turk [68]. However, recent research [69,70] reports a semi-supervised deep learning approach using word-embedding [71]. Word-embedding learns dense vector representations of words and phrases while automatically preserving the semantic relationships that exist in the text corpus by incorporating such relations into the vector space of the word-embedding. This enables the use of linear algebra to capture different semantic relationships within word-vectors in the word-embedding. The famous example in [72] shows that the vector arithmetic of word vectors 'King -Man + Woman' results a word vector similar to the word vector of 'Queen'.
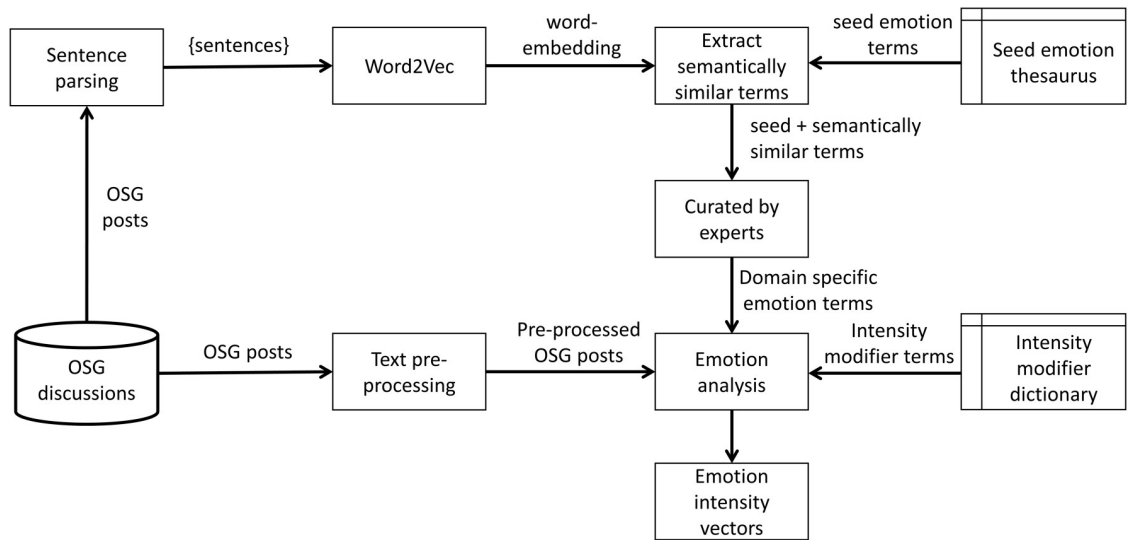
Fig 3. The proposed technique for emotion extraction.

Developing such a word-embedding using OSG discussions enables to capture terms used by the OSG users that are semantically similar to the seed emotional terms. We have developed a word-embedding from a large text corpus which contained a total of 4,795,428 OSG posts. This corpus was pre-processed to remove URLs, convert to lower case and then separated into sentences using the Punkt sentence tokenizer [73] available in python NLTK library [74], which has shown state-of-the-art performance when compared to other sentence tokenizers with over 90% accuracy on user generated content [75]. This tokenization has resulted in 36,222,536 sentences. This text corpus was used to train a 200 dimensional word-embedding using Word2Vec technique with skip-gram model [71] and negative-sampling [76]. We utilised the python genism [77] library for this implementation. The resulting word-embedding contains 312,196 unique terms (words and phrases).

Following the trained word-embedding, top 25 most similar terms for each seed term in the emotion thesaurus was identified using a nearest neighbour search in the embedding space using Cosine similarity. These identified terms are semantically similar terms to the seed emotion terms, in which some of the terms have the same emotional sense of the seed term while some others may not. For example, the top five nearest neighbours of *sorrowful* are *sadness*, *sincerity*, *joyful*, and *deeply saddened*, in which *joyful* is semantically similar but has the opposite emotional sense. Therefore, a further empirical validation was also conducted. The third column of Table 3 presents a sample of emotional terms captured using the above technique.

Intensity modifier terms are a set of terms that increase or decrease the intensity of the emotional term. For example, the term 'very' increases the intensity of the emotion 'good' when used together, whereas, the term 'kind of' decreases the intensity of the emotion 'okay' when used together. Moreover, some terms completely negate the emotions e.g., 'not okay' negates the emotion expressed by 'okay'. A thesaurus of such terms are often used in rule based sentiment analysis tools such as SentiStrength [78] and VADER [79] to improve the accuracy of the sentiment score. In this work, we have used the intensity modifier term thesaurus used in VADER [79]. S1 Fig provides the algorithm for calculating emotions vector E_P of a given OSG post P.

In summary, as explicated above, PRIME functions in seven stages S1-S7 to transform OSG discussions from unstructured text discussions in the everyday language into multi-granular,

Table 3. Emotion categories and a sample of representative terms used for each emotion.

Emotion	Emotion terms from thesaurus	Emotion terms extracted by proposed technique
Positive emotions		
Happy	happy, great, joyous, glad, delighted	fab, chuffed, terrific, great news, looking forward, heart warming, uplifting, upbeat
Good	good, pleased, comfortable, relaxed, content	comfy, nice, chill, chipper, ok, okay, clear headed, cool
Alive	alive, playful, energetic, spirited, animated	chatty, perky, sociable, vibrant, vivacious, witty, easy going, peppy
Love	love, attracted, warm, passionate, affectionate	romantic, cuddly, compassionate, intimate, adore, supportive, caring
Positive	positive, eager, keen, bold, brave	smart, ambitious, proactive, cynical, insistent, willing, upbeat
Open	open, understanding, accepting, satisfied, receptive	open minded, empathetic, cooperative, accommodating, approachable, forgiving, attuned, rational
Interested	interested, fascinated, inquisitive, curious, intrigued	keen, impressed, cautious, leery, eager, intuitive, savvy, thoughtful
Strong	strong, certain, dynamic, sure, tenacious	resilient, independent, adamant, fierce, self reliant, decisive, fighter, pragmatic
Negative emotions		
Sad	sad, tearful, grief, sorrowful, grief	heart break, teary, lonely, weepy, crying, despairing, hurtful
Afraid	afraid, fearful, terrified, panic, worry	petrified, freaking out, apprehensive, dread, obsess, fret, nervous wreck
Hurt	hurt, deprived, pained, dejected, agonised	traumatised, bruised, shattered, ached, exhausted, cramped, numb, fatigued, strained
Angry	angry, annoy, provoke, aggressive, enraged	agitated, hostile, pissed off, argumentative, aggressive, rude, paranoid, ticked off, lashing out
Depressed	depressed, disappointed, miserable, despair, powerless	despondent, distraught, suicidal, unloved, worthless, emotionally drained, snappy
Helpless	helpless, incapable, alone, vulnerable, fatigued	insecure, tired, hopeless, powerless, defeated, overwhelmed, listless, incapacitated
Confused	confused, upset, doubtful, uncertain, hesitant	unsure, perplexed, wary, leery freaked out, iffy, bummed, taken aback
Indifferent	indifferent, insensitive, dull, reserved, lifeless	grumpy, apathetic, blunt, ignorant, emotionless, callous, crass, standoffish

https://doi.org/10.1371/journal.pone.0205855.t003

multidimensional information individualised by the patient to analyse and aggregate 'real life' patient reported outcomes.

Statistical analysis

The differences in variables between the groups were analysed using the Chi-Squared test (categorical) and 2-sided student's t-test (means). The differences in side effects at selected time points were analysed using the chi-squared test, while the student's t-test was used to compare differences between average emotion scores. $P < 0.05$ was used for statistical significance. Analysis was performed using SAS software version 9.4.

Results

PRIME was applied on ten high volume active OSG focused on PCa discussions, a dataset contains 609,960 conversations from 22,233 patients, comprising a text corpus of 93,606,581 word tokens. Following results are based on the inclusion criteria for this study; patients who self-disclosed their chosen PCa treatment and discussed the decision-making process that led to the selection.

Fig 4(a)–4(c) present the composition of each group in terms of volume, age, grading of cancer (using Gleason score) and modality of treatment. The paternalistic group is significantly smaller with a uniform distribution of age, whereas Autonomous, Shared groups are of comparable volume with approximately normal distribution of age. A high percentage of patients had Gleason<7 PCa and chose surgery as the treatment option. Fig 4(d)–4(f) represent the monthly trajectory of patient activity on OSG, three months before (-3) to 12 months following the decision. The timing of each decision is extracted by PRIME using an incremental machine learning technique [80]. In Fig 4(d), the noticeable peak of patient activity across all three groups during the period of decision-making (Paternalistic = 420, Autonomous = 3883, Shared = 2154), is an indication of active information seeking by all groups. Paternalistic and Autonomous groups reduce activity soon afterward, but the Shared group
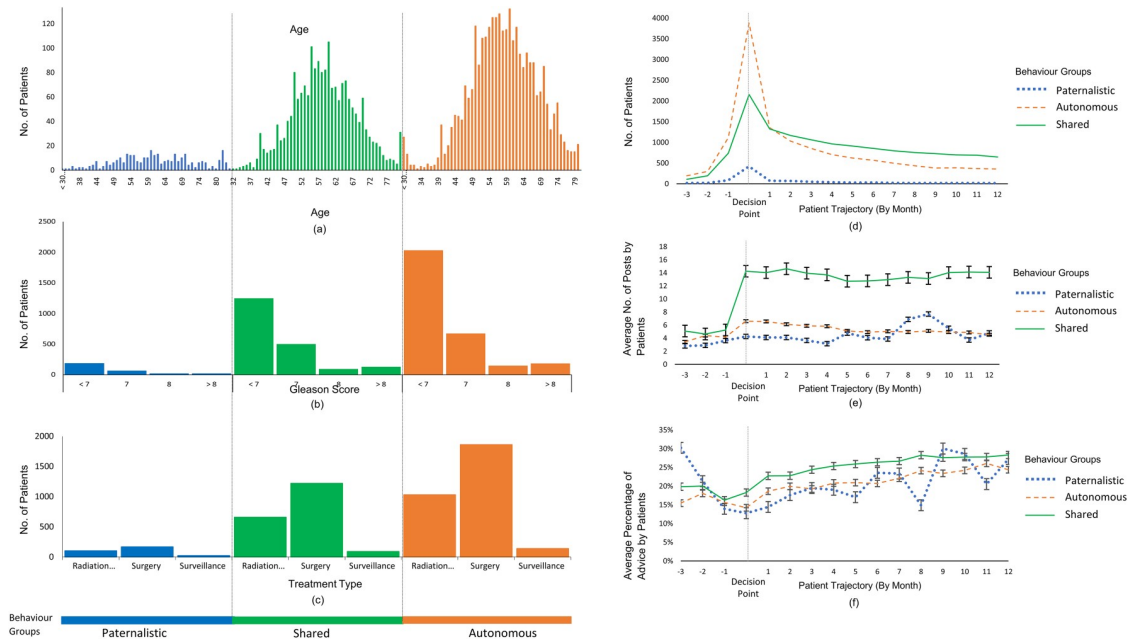
Fig 4. PRIME extracts multiple modalities of information individually for each patient from OSG discussions. This individual information is grouped into corresponding decision-making behaviour; Paternalistic, Autonomous and Shared groups by (a) patient age, (b) Gleason score and (c) treatment type. PRIME further generates aggregated trajectories for multiple temporal aspects of each decision-making behaviour group, before and after the decision, (d) number of patients engaged in OSG discussions, (e) average number of posts and (f) average percentage of posts containing advice by each group. The decision point is indicated by dotted vertical line in d-f.

consistently participate in OSG discussions throughout the 12 months, Fig 4(e). PRIME can automatically distinguish between forum posts providing advice from those seeking answers/ sharing experiences. Fig 4(f) reports the trajectory for percentage of advice posts by each group. Overall, the percentage of advice posts is lowest for the first month even though the average number of posts are highest. This number gradually increases across 12 months and interestingly, hitherto dormant Paternalistic group actively partakes in providing advice. A general trend observed in Fig 4 is that Autonomous group consistently participates over the given time period whereas Shared group demonstrates an increased interest in receiving and sharing following their treatment decision.

Fig 5 illustrates the diversity of decision factors, ranging from clinical skills to financial concerns. It can be observed that 'doctor experience' is most influential (65%) across all three behaviour groups. Shared Group discusses all decision factors significantly more (p<0.001) than the other two groups. In treatment options, Surgery and Surveillance Groups take into consideration most factors whereas Radiation Group is more concerned about radiation oncologist (21.66%), bowel symptoms (2.04%) and financial concerns (7.94%). Age distributions (Fig 3(c)) is also diverse with fringe groups (<40 and >70 less concerned about 'doctor experience' than other groups and age group 51–60 is relatively more influenced by clinical factors than personal factors.

Fig 6 presents a comprehensive analysis of distinct patient emotions, expressed over time— from pre-decision to recovery. Besides slight peaks at the decision point, the Shared group is mostly consistent in expressing negative and positive emotion. The paternalistic group is significantly more expressive, with far less positive emotions (aggregate of -7.48) and strongly expressed emotions related to 'depressed'. Interestingly, this group expressed less negative

	C	P	CP	Sub Total	p-value
Doctor skill discussed	200 (47.62)	2162 (55.68)	1837 (85.28)	4199 (65.03)	0
Surgeon mentioned	219 (52.14)	1047 (26.96)	1730 (80.32)	2996 (46.4)	0
Radiation Oncologist mentioned	51 (12.14)	186 (4.79)	649 (30.13)	886 (13.72)	0
GP mentioned	12 (2.86)	51 (1.31)	165 (7.66)	228 (3.53)	0
Medical Indication	8 (1.9)	187 (4.82)	360 (16.71)	555 (8.6)	0
Nerve Sparing mentioned	20 (4.76)	299 (7.7)	490 (22.75)	809 (12.53)	0
Young age	6 (1.43)	110 (2.83)	265 (12.3)	381 (5.9)	0
Bowel Symptoms	0 (0)	18 (0.46)	42 (1.95)	60 (0.93)	0
Urinary incontinence	20 (4.76)	301 (7.75)	507 (23.54)	828 (12.82)	0
Erectile dysfunction	13 (3.1)	205 (5.28)	482 (22.38)	700 (10.84)	0
Recovery Time	3 (0.71)	77 (1.98)	184 (8.54)	264 (4.09)	0
Best Cancer Control	22 (5.24)	331 (8.52)	587 (27.25)	940 (14.56)	0
Financial Concerns	10 (2.38)	136 (3.5)	386 (17.92)	532 (8.24)	0
Less invasive	2 (0.48)	42 (1.08)	105 (4.87)	149 (2.31)	0

(a)

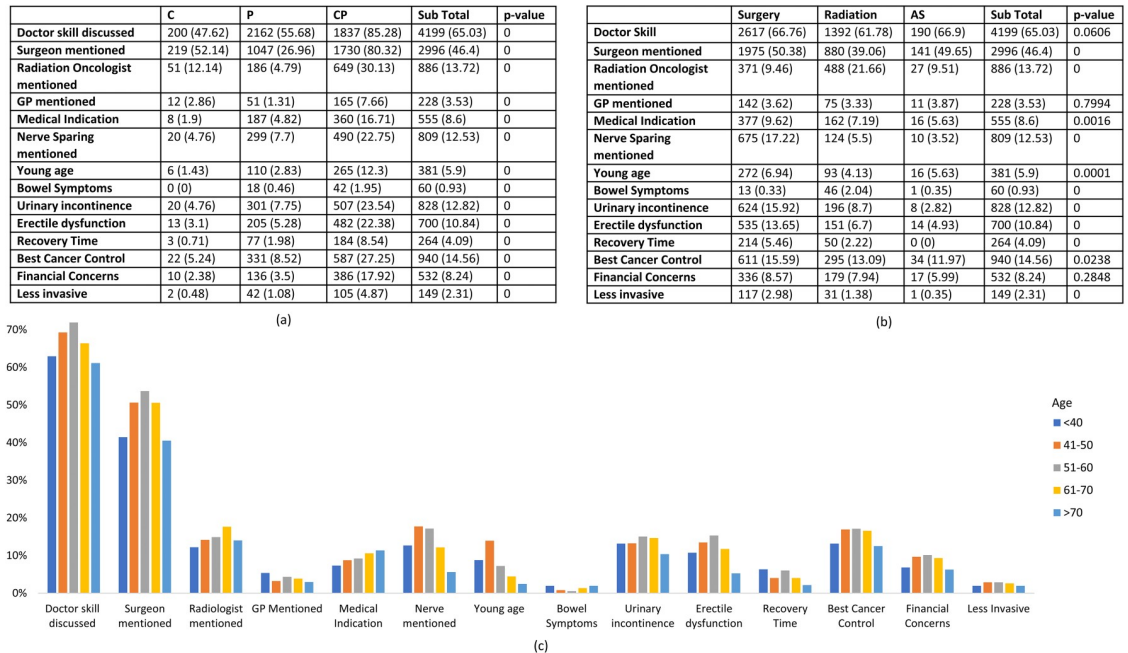|	Surgery	Radiation	AS	Sub Total	p-value
Doctor Skill	2617 (66.76)	1392 (61.78)	190 (66.9)	4199 (65.03)	0.0606
Surgeon mentioned	1975 (50.38)	880 (39.06)	141 (49.65)	2996 (46.4)	0
Radiation Oncologist mentioned	371 (9.46)	488 (21.66)	27 (9.51)	886 (13.72)	0
GP mentioned	142 (3.62)	75 (3.33)	11 (3.87)	228 (3.53)	0.7994
Medical Indication	377 (9.62)	162 (7.19)	16 (5.63)	555 (8.6)	0.0016
Nerve Sparing mentioned	675 (17.22)	124 (5.5)	10 (3.52)	809 (12.53)	0
Young age	272 (6.94)	93 (4.13)	16 (5.63)	381 (5.9)	0.0001
Bowel Symptoms	13 (0.33)	46 (2.04)	1 (0.35)	60 (0.93)	0
Urinary incontinence	624 (15.92)	196 (8.7)	8 (2.82)	828 (12.82)	0
Erectile dysfunction	535 (13.65)	151 (6.7)	14 (4.93)	700 (10.84)	0
Recovery Time	214 (5.46)	50 (2.22)	0 (0)	264 (4.09)	0
Best Cancer Control	611 (15.59)	295 (13.09)	34 (11.97)	940 (14.56)	0.0238
Financial Concerns	336 (8.57)	179 (7.94)	17 (5.99)	532 (8.24)	0.2848
Less invasive	117 (2.98)	31 (1.38)	1 (0.35)	149 (2.31)	0

(b)

(c)

Fig 5. Key decision factors for PCa patients, in terms of (a) decision-making behaviour groups (b) treatment modality and (c) age groups. In (a) and (b), p-values were calculated for statistical significance.

https://doi.org/10.1371/journal.pone.0205855.g005

emotions during 8–9 months with an immediate increase in 10–11 months. The Autonomous group demonstrate a similar pattern, less remarkably, during 8–9 months. Age group <40 consistently express above average positive emotions, with 'love' most expressed. On negative emotions, <40 and >70 groups are consistently above average with Surveillance group significantly below average.

Fig 7 reflects on side effects, with a higher percentage of Shared group reporting all side effects than the other two groups. Shared and Autonomous groups are initially affected by significant urinary side effects (Fig 7(a)) which gradual decline over time, in contrast to the Paternalistic Group which has fluctuations over time. Shared and Autonomous groups are consistently affected by sexual side effects (Fig 7(b)) while Paternalistic group shows an increase over time, reaching a level equivalent to that of the Shared group by month 12. Bowel side effects are least mentioned as the numbers opting for radiation treatment are significantly less. Fig 7(d), Paternalistic group exhibit a peak in other side effects (mainly, infections and bleeding) during the tenth month. As anticipated, sexual and urinary side effects are strongly expressed by younger age groups whereas bowel and other side effects are more consistent across all age groups, Fig 7(f).

Discussion

Results generated by our PRIME framework strongly correlate with three major patient behaviour groups (Autonomous, Paternalistic and Shared) [54]. All groups actively sought information on OSG: The Shared group provided consistent, prolonged interactions, sharing their positive and negative emotions, experiences and advice, while the Paternalistic group were more expressive, especially with negative emotions but contributed to the OSG with advice many months post-treatment. The Autonomous group only sought advice and contributed minimally to conversations on OSG. These ecosystem-like interactions indicate the self-
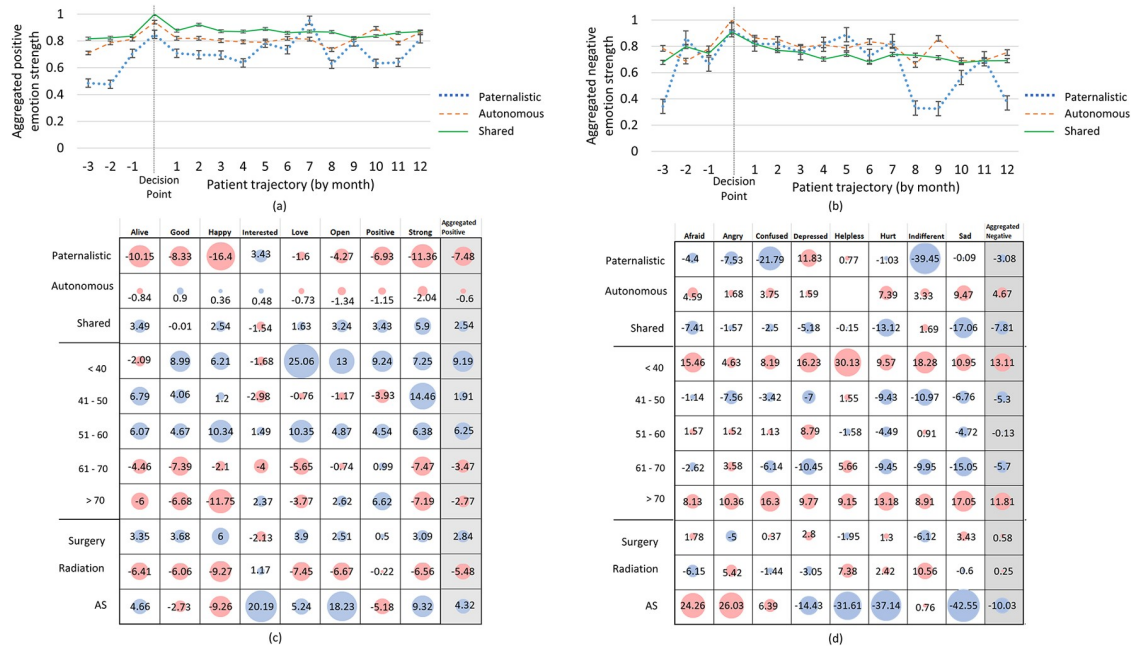
Fig 6. OSG provide insights into emotional journeys of patients making an intimate decision of selecting a treatment type following the diagnosis of cancer, (a) aggregated positive emotion over time, (b) aggregated negative emotion over time as well as relative strength of each distinct emotion by behaviour group, age group, treatment option, for positive emotions (c) and negative emotions (d).

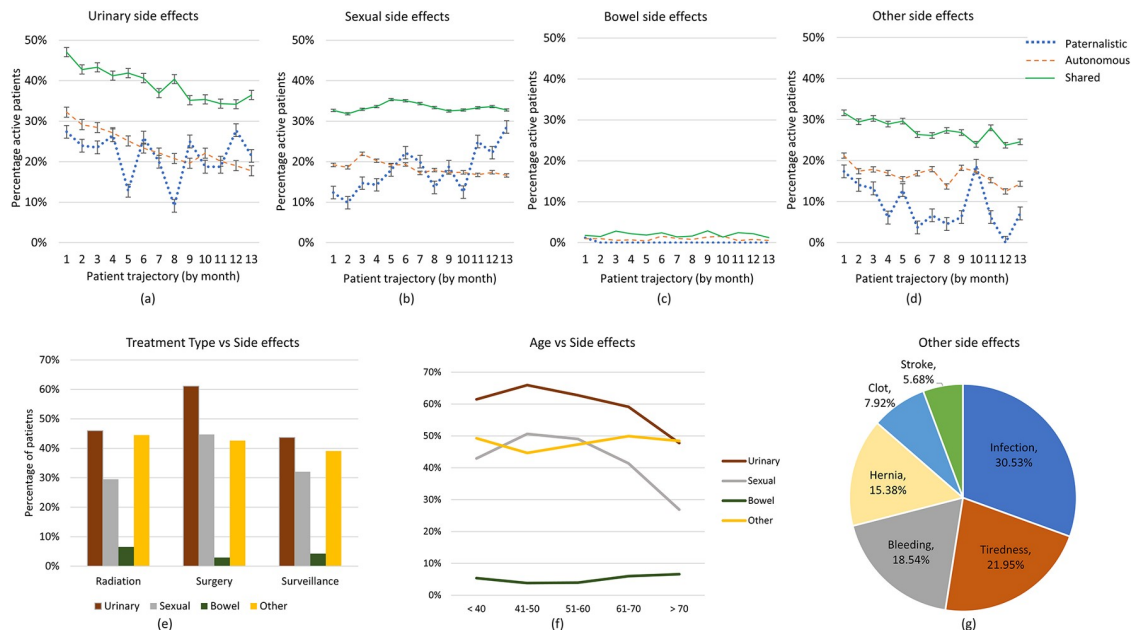https://doi.org/10.1371/journal.pone.0205855.g006

Fig 7. PRIME extracts and associates side effects mentioned and discussed on OSG with all other aspects of extracted patient information to generate trajectories for behaviour groups (a) urinary side effects, (b) sexual side effects, (c) bowel side effects and (d) other side effects. Further illustrated (e) side effects by treatment options, (f) side effects by age groups and (g) the general composition of other side effects.

https://doi.org/10.1371/journal.pone.0205855.g007

sufficient nature of OSG where patient voices are prominently and equally represented. Thereby, it is timely and relevant for primary care providers to accept OSG as an adjunct to cancer care and consider participating in OSG through artificial intelligence enabled optimised moderation and streamlined intervention.

In healthcare decision-making, patients continue to be an unheard, often forgotten voice [81,82]. Despite stringent efforts to advance the paradigm of patient-centred care [83,84], the importance of shared decision-making continues to be overlooked [85]. It is argued that patients should be provided necessary tools to gather information, know their decision options, scenarios and consequences for shared decision-making to be effective [86]. The significance of emotional support that allows patients to freely express values and preferences and ask questions without clinician obstruction is also highlighted [87]. The proliferation of OSG is a clear indication that patients and carers are bridging this gap by seeking (and providing) this service extraneous to healthcare providers and institutions. Further, OSG provide information, decision options and emotional support with the added advantage of a geographically dispersed community of individuals who are undergoing/have undergone similar circumstances [88,89].

Besides fulfilling the essential role of decision support for patient-centred care, OSG make a further paramount contribution as a medium for post-decision conversations on information exchange and emotional support. This is seen to be instrumental in addressing the 'out of sight out of mind' dilemma that arises due to periodic and/or occasional clinician consultations during the recovery phase. Patients who have undergone similar treatment are willing to share their experience, offer advice and emotional support during this crucial recovery period. Although patients with similar experiences provide each other support, OSG are peer to peer and unregulated which can be challenging for optimal healthcare. Therefore, healthcare providers must identify specific patient needs communicated on OSG, in order to optimise delivery of care and ensure that patients don't extensively depend on their peers for healthcare advice. However, healthcare providers and institutions are progressively limited in their scope of reach and service, due to increased demand, financial constraints, resource limitations and employee turnover [90,91]. It is difficult to transition from disease-centred to patient-centred healthcare delivery in such volatile settings. With increased utilisation of OSG and the increasing presence of social media empowered patients, the medical support network for cancer care must evolve to integrate these platforms in order to provide optimal and individualised care that is clinically appropriate for patients with cancer. As explicated in this study, the PRIME framework provides significant evidence supporting the need for an optimised, cost-effective, and integrated platform for patient focused healthcare delivery.

Conclusion

In summary, PRIME is an artificial intelligence based analytics framework for supporting social media empowered patients. It can be used for automatic aggregation and investigation of patient decision-making behaviours, decision factors, social interaction trajectory pre-/post- decision-making as well as positive and negative emotion trajectory pre-/post- decision-making. We have demonstrated these novel functionalities on patients with prostate cancer, from diagnosis to treatment and recovery. PRIME demonstrates, quantitatively, how patients in OSG evolve from being information seekers to providers, over time as they progress from treatment to recovery. Automatic aggregation and profiling patients, using machine learning and natural language processing, based on their decision-making characteristics, side effects and emotions indicate the practical value of PRIME towards informed decision-making on personalized healthcare delivery and the development of policy

guidelines for primary care moderation and interventions, by clinicians, psychologists and other cancer care providers.

Supporting information

S1 Fig. Algorithm for determining the 16-dimensional emotion vector E_P of a given OSG post P.
(PDF)

Author Contributions

Conceptualization: Daswin De Silva, Weranja Ranasinghe, Tharindu Bandaragoda, Achini Adikari, Damminda Alahakoon, Nathan Lawrentschuk, Raj Persad, Richard Gray, Damien Bolton.

Data curation: Daswin De Silva, Tharindu Bandaragoda, Achini Adikari, Lahiru Iddamalgoda.

Formal analysis: Daswin De Silva, Weranja Ranasinghe, Tharindu Bandaragoda, Achini Adikari, Nishan Mills, Nathan Lawrentschuk, Raj Persad, Richard Gray, Damien Bolton.

Investigation: Daswin De Silva, Weranja Ranasinghe, Tharindu Bandaragoda, Achini Adikari, Nishan Mills, Lahiru Iddamalgoda, Damminda Alahakoon, Nathan Lawrentschuk, Raj Persad, Evgeny Osipov, Richard Gray, Damien Bolton.

Methodology: Daswin De Silva, Weranja Ranasinghe, Tharindu Bandaragoda, Achini Adikari, Nishan Mills, Lahiru Iddamalgoda, Damminda Alahakoon, Nathan Lawrentschuk, Raj Persad, Evgeny Osipov, Richard Gray, Damien Bolton.

Project administration: Daswin De Silva, Damminda Alahakoon.

Software: Evgeny Osipov.

Supervision: Daswin De Silva, Weranja Ranasinghe, Damminda Alahakoon, Nathan Lawrentschuk, Raj Persad, Richard Gray, Damien Bolton.

Validation: Daswin De Silva, Tharindu Bandaragoda, Achini Adikari, Lahiru Iddamalgoda, Damminda Alahakoon, Evgeny Osipov.

Visualization: Daswin De Silva, Tharindu Bandaragoda, Achini Adikari, Nishan Mills, Lahiru Iddamalgoda, Evgeny Osipov.

Writing – original draft: Daswin De Silva, Weranja Ranasinghe, Tharindu Bandaragoda, Achini Adikari.

Writing – review & editing: Daswin De Silva, Weranja Ranasinghe, Tharindu Bandaragoda, Achini Adikari, Nishan Mills, Damminda Alahakoon, Nathan Lawrentschuk, Raj Persad, Evgeny Osipov, Richard Gray, Damien Bolton.

References

1. Fox S. The Social Life of Health Information. Pew Internet \& American Life Project Washington, DC; 2011.

2. Westerman D, Spence PR, Van Der Heide B. Social Media as Information Source: Recency of Updates and Credibility of Information. J Comput Commun. Oxford University Press; 2014; 19: 171–183. https://doi.org/10.1111/jcc4.12041

3. Lee K, Hoti K, Hughes JD, Emmerton LM. Interventions to assist health consumers to find reliable online health information: A comprehensive review. Manchikanti L, editor. PLoS One. Public Library of Science; 2014; 9: e94186. https://doi.org/10.1371/journal.pone.0094186 PMID: 24710348

4. Wright K. Social support within an on-line cancer community: An assessment of emotional support, perceptions of advantages and disadvantages, and motives for using the community from a communication perspective. J Appl Commun Res. 2002; 30: 195–209. https://doi.org/10.1080/00909880216586

5. Barak A, Boniel-Nissim M, Suler J. Fostering empowerment in online support groups. Comput Human Behav. 2008; 24: 1867–1883. https://doi.org/10.1016/j.chb.2008.02.004

6. van Uden-Kraan CF, Drossaert CHC, Taal E, Shaw BR, Seydel ER, van de Laar MAFJ. Empowering Processes and Outcomes of Participation in Online Support Groups for Patients With Breast Cancer, Arthritis, or Fibromyalgia. Qual Health Res. 2008; 18: 405–417. https://doi.org/10.1177/1049732307313429 PMID: 18235163

7. Mo PKH, Coulson NS. Exploring the Communication of Social Support within Virtual Communities: A Content Analysis of Messages Posted to an Online HIV/AIDS Support Group. CyberPsychology Behav. 2008; 11: 371–374. https://doi.org/10.1089/cpb.2007.0118 PMID: 18537512

8. Evans M, Donelle L, Hume-Loveland L. Social support and online postpartum depression discussion groups: A content analysis. Patient Educ Couns. 2012; 87: 405–410. https://doi.org/10.1016/j.pec.2011.09.011 PMID: 22019021

9. Naslund JA, Grande SW, Aschbrenner KA, Elwyn G. Naturally occurring peer support through social media: The experiences of individuals with severe mental illness using you tube. De Haan L, editor. PLoS One. Public Library of Science; 2014; 9: e110171. https://doi.org/10.1371/journal.pone.0110171 PMID: 25333470

10. Jayles B, Kim H, Escobedo R, Cezera S, Blanchet A, Kameda T, et al. How social information can improve estimation accuracy in human groups. Proc Natl Acad Sci. 2017; 114: 201703695. https://doi.org/10.1073/pnas.1703695114 PMID: 29118142

11. Mishra MV., Bennett M, Vincent A, Lee OT, Lallas CD, Trabulsi EJ, et al. Identifying barriers to patient acceptance of active surveillance: content analysis of online patient communications. Sarkar IN, editor. PLoS One. Public Library of Science; 2013; 8: e68563. https://doi.org/10.1371/journal.pone.0068563 PMID: 24039699

12. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA. American Medical Association; 2013; 309: 1351–1352. https://doi.org/10.1001/jama.2013.393 PMID: 23549579

13. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Aff. 2014; 33: 1123–1131. https://doi.org/10.1377/hlthaff.2014.0041 PMID: 25006137

14. Obermeyer Z, Emanuel EJ. Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. Massachusetts Medical Society; 2016; 375: 1216–1219. https://doi.org/10.1056/NEJMp1606181 PMID: 27682033

15. Shah ND, Steyerberg EW, DM K. Big data and predictive analytics: Recalibrating expectations. JAMA. 2018;

16. Shaikh AR, Butte AJ, Schully SD, Dalton WS, Khoury MJ, Hesse BW. Collaborative biomedicine in the age of big data: The case of cancer. J Med Internet Res. Journal of Medical Internet Research; 2014; 16: e101. https://doi.org/10.2196/jmir.2496 PMID: 24711045

17. Badawi O, Brennan T, Celi LA, Feng M, Ghassemi M, Ippolito A, et al. Making big data useful for health care: A summary of the inaugural MIT critical data conference. Journal of Medical Internet Research. JMIR Publications Inc.; 2014. p. e22. https://doi.org/10.2196/medinform.3447 PMID: 25600172

18. LeCun YA, Bengio Y, Hinton GE. Deep learning. Nature. Nature Publishing Group; 2015; 521: 436–444. https://doi.org/10.1038/nature14539 PMID: 26017442

19. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep Learning for Health Informatics. IEEE J Biomed Heal Informatics. 2017; 21: 4–21. https://doi.org/10.1109/JBHI.2016.2636665 PMID: 28055930

20. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform. 2017; https://doi.org/10.1093/bib/bbx044 PMID: 28481991

21. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep. Nature Publishing Group; 2016; 6: 26094. https://doi.org/10.1038/srep26094 PMID: 27185194

22. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? Journal of Biomedical Informatics. Academic Press; 2009. pp. 760–772. https://doi.org/10.1016/j.jbi.2009.08.007 PMID: 19683066

23. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA—J Am Med Assoc. American Medical Association; 2011; 306: 848–855. https://doi.org/10.1001/jama.2011.1204 PMID: 21862746

24. Yim W, Yetisgen M, Harris WP, Kwan SW. Natural Language Processing in Oncology: A Review. JAMA Oncol. 2016; 2: 1–8. https://doi.org/10.1001/jamaoncol.2016.0213 PMID: 27124593

25. Rumshisky A, Ghassemi M, Naumann T, Szolovits P, Castro VM, McCoy TH, et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. Transl Psychiatry. Nature Publishing Group; 2016; 6: e921. https://doi.org/10.1038/tp.2015.182 PMID: 27754482

26. Lu Y, Zhang P, Liu J, Li J, Deng S. Health-Related Hot Topic Detection in Online Communities Using Text Clustering. Bullen C, editor. PLoS One. Public Library of Science; 2013; 8: e56221. https://doi.org/10.1371/journal.pone.0056221 PMID: 23457530

27. Ranasinghe W, Bandaragoda T, De Silva D, Alahakoon D. A novel framework for automated, intelligent extraction and analysis of online support group discussions for cancer related outcomes. BJU Int. 2017; 120: 59–61. https://doi.org/10.1111/bju.14036 PMID: 29058379

28. Bandaragoda TR, De Silva D, Alahakoon D, Ranasinghe W, Bolton D. Text mining for personalised knowledge extraction from online support groups. J Assoc Inf Sci Technol. 2018;

29. Bandaragoda T, Ranasinghe W, Adikari A, de Silva D, Lawrentschuk N, Alahakoon D, et al. The Patient-Reported Information Multidimensional Exploration (PRIME) Framework for Investigating Emotions and Other Factors of Prostate Cancer Patients with Low Intermediate Risk Based on Online Cancer Support Group Discussions. Ann Surg Oncol. Springer International Publishing; 2018; 25: 1737–1745. https://doi.org/10.1245/s10434-018-6372-2 PMID: 29468607

30. Ranasinghe W, de Silva D, Bandaragoda T, Adikari A, Alahakoon D, Persad R, et al. Robotic-assisted vs. open radical prostatectomy: A machine learning framework for intelligent analysis of patient-reported outcomes from online cancer support groups. Urol Oncol Semin Orig Investig. Elsevier; 2018; https://doi.org/10.1016/j.urolonc.2018.08.012 PMID: 30236854

31. Welch HG, Albertsen PC. Prostate Cancer Diagnosis and Treatment After the Introduction of Prostate-Specific Antigen Screening: 1986–2005. JNCI J Natl Cancer Inst. Oxford University Press; 2009; 101: 1325–1329. https://doi.org/10.1093/jnci/djp278 PMID: 19720969

32. American Cancer Society. Am Cancer Soc. 2017;

33. Worldwide cancer statistics. In: Cancer Research UK. May 2015.

34. Harrison JD, Young JM, Price MA, Butow PN, Solomon MJ. What are the unmet supportive care needs of people with cancer? A systematic review. Support Care Cancer. 2009; 17: 1117–1128. https://doi.org/10.1007/s00520-009-0615-5 PMID: 19319577

35. Flood AB, Wennberg JE, Nease RF Jr, Fowler FJ Jr, Ding J, Hynes LM. The importance of patient preference in the decision to screen for prostate cancer. Prostate Patient Outcomes Research Team. J Gen Intern Med. 1996; 11: 342–349. https://doi.org/10.1007/BF02600045 PMID: 8803740

36. Blank TO, Schmidt SD, Vangsness SA, Monteiro AK, Santagata PV. Differences among breast and prostate cancer online support groups. Comput Human Behav. Elsevier Ltd; 2010; 26: 1400–1404. https://doi.org/10.1016/j.chb.2010.04.016

37. Coulson NS, Buchanan H, Aubeeluck A. Social support in cyberspace: A content analysis of communication within a Huntington's disease online support group. Patient Educ Couns. 2007; 68: 173–178. https://doi.org/10.1016/j.pec.2007.06.002 PMID: 17629440

38. Coulson NS. Receiving Social Support Online: An Analysis of a Computer-Mediated Support Group for Individuals Living with Irritable Bowel Syndrome. CyberPsychology Behav. 2005; 8: 580–584. https://doi.org/10.1089/cpb.2005.8.580 PMID: 16332169

39. Bar-Lev S. "We are here to give you emotional support": Performing emotions in an online HIV/AIDS support group. Qual Health Res. 2008; 18: 509–521. https://doi.org/10.1177/1049732307311680 PMID: 18192435

40. Lieberman MA, Goldstein BA. Not all negative emotions are equal: The role of emotional expression in online support groups for women with breast cancer. Psychooncology. John Wiley & Sons, Ltd.; 2006; 15: 160–168. https://doi.org/10.1002/pon.932 PMID: 15880627

41. Ruthven I, Buchanan S, Jardine C. Isolated, overwhelmed, and worried: Young first-time mothers asking for information and support online. Journal of the Association for Information Science and Technology. 2018.

42. Huber J, Ihrig A, Peters T, Huber CG, Kessler A, Hadaschik B, et al. Decision-making in localized prostate cancer: Lessons learned from an online support group. BJU Int. 2011; 107: 1570–1575. https://doi.org/10.1111/j.1464-410X.2010.09859.x PMID: 21105988

43. Chen AT, Zhu S-H, Conway M. What Online Communities Can Tell Us About Electronic Cigarettes and Hookah Use: A Study Using Text Mining and Visualization Techniques. J Med Internet Res. JMIR Publications Inc.; 2015; 17: e220. https://doi.org/10.2196/jmir.4517 PMID: 26420469

44. Chen AT. Exploring online support spaces: Using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient Educ Couns. Elsevier Ireland Ltd; 2012; 87: 250–257. https://doi.org/10.1016/j.pec.2011.08.017 PMID: 21930359

45. Lu Y, Zhang P, Liu J, Li J, Deng S. Health-Related Hot Topic Detection in Online Communities Using Text Clustering. PLoS One. 2013; 8: 1–9. https://doi.org/10.1371/journal.pone.0056221 PMID: 23457530

46. Wang YC, Kraut RE, Levine JM. Eliciting and receiving online support: Using computer-aided content analysis to examine the dynamics of online social support. J Med Internet Res. Journal of Medical Internet Research; 2015; 17: e99. https://doi.org/10.2196/jmir.3558 PMID: 25896033

47. Tausczik YR, Pennebaker JW. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. J Lang Soc Psychol. 2010; 29: 24–54. https://doi.org/10.1177/0261927X09351676

48. Shim M, Cappella JN, Han JY. How Does Insightful and Emotional Disclosure Bring Potential Health Benefits? Study Based on Online Support Groups for Women With Breast Cancer. J Commun. NIH Public Access; 2011; 61: 432–454. https://doi.org/10.1111/j.1460-2466.2011.01555.x PMID: 25568496

49. Alpers GW, Winzelberg AJ, Classen C, Roberts H, Dev P, Koopman C, et al. Evaluation of computerized text analysis in an Internet breast cancer support group. Comput Human Behav. Pergamon; 2005; 21: 361–376. https://doi.org/10.1016/j.chb.2004.02.008

50. Yates A, Cohan A, Goharian N. Depression and Self-Harm Risk Assessment in Online Forums. 2017 Conference on Empirical Methods in Natural Language Processing. 2017. pp. 2968–2978.

51. Donovan JL, Hamdy FC, Lane JA, Mason M, Metcalfe C, Walsh E, et al. Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. N Engl J Med. 2016; 375: 1425–1437. https://doi.org/10.1056/NEJMoa1606221 PMID: 27626365

52. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. Oxford University Press; 2004; 32: 267D–270. https://doi.org/10.1093/nar/gkh061 PMID: 14681409

53. Aronson AR, Lang F-M, Aronson A, Aronson A, Rindflesch T, Browne A, et al. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc. The Oxford University Press; 2010; 17: 229–36. https://doi.org/10.1136/jamia.2009.002733 PMID: 20442139

54. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: Revisiting the shared treatment decision-making model. Soc Sci Med. 1999; 49: 651–661. https://doi.org/10.1016/S0277-9536(99)00145-8 PMID: 10452420

55. Hu X, Bell RA, Kravitz RL, Orrange S. The Prepared Patient: Information Seeking of Online Support Group Members Before Their Medical Appointments. J Health Commun. 2012; 17: 960–978. https://doi.org/10.1080/10810730.2011.650828 PMID: 22574697

56. Ihrig A, Keller M, Hartmann M, Debus J, Pfitzenmaier J, Hadaschik B, et al. Treatment decision-making in localized prostate cancer: Why patients chose either radical prostatectomy or external beam radiation therapy. BJU Int. Wiley/Blackwell (10.1111); 2011; 108: 1274–1278. https://doi.org/10.1111/j.1464-410X.2011.10082.x PMID: 21410634

57. Berry DL, Ellis WJ, Woods NF, Schwien C, Mullen KH, Yang C. Treatment decision-making by men with localized prostate cancer: The influence of personal factors. Urol Oncol Semin Orig Investig. Elsevier; 2003; 21: 93–100. https://doi.org/10.1016/S1078-1439(02)00209-0

58. Gwede CK, Pow-Sang J, Seigne J, Heysek R, Helal M, Shade K, et al. Treatment decision-making strategies and influences in patients with localized prostate carcinoma. Cancer. Wiley-Blackwell; 2005; 104: 1381–1390. https://doi.org/10.1002/cncr.21330 PMID: 16080181

59. Huber J, Maatz P, Muck T, Keck B, Friederich HC, Herzog W, et al. The effect of an online support group on patients' treatment decisions for localized prostate cancer: An online survey. Urol Oncol Semin Orig Investig. Elsevier; 2017; 35: 37.e19–37.e28. https://doi.org/10.1016/j.urolonc.2016.09.010 PMID: 27810256

60. Gupta S, Maclean DL, Heer J, Manning CD. Induced lexico-syntactic patterns improve information extraction from online medical forums. J Am Med Informatics Assoc. 2014; 21: 902–909. https://doi.org/10.1136/amiajnl-2014-002669 PMID: 24970840

61. Oh S. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. J Am Soc Inf Sci Technol. Wiley Subscription Services, Inc., A Wiley Company; 2012; 63: 543–557. https://doi.org/10.1002/asi.21676

62. Bar-Lev S. "We are here to give you emotional support": performing emotions in an online HIV/AIDS support group. Qual Health Res. Sage PublicationsSage CA: Los Angeles, CA; 2008; 18: 509–521. https://doi.org/10.1177/1049732307311680 PMID: 18192435

63. Russell JA. A circumplex model of affect. J Psychol Soc Psychol. 1980; 39: 1161.

64. Plutchik R. The Emotions. University Press of America; 1991.

65. Mohammad SM. Sentiment Analysis: Detecting Valence, Emotions, and Other Affected States from Text. Emotion Measurement. 2016. pp. 201–237. https://doi.org/10.1016/B978-0-08-100508-8.00009-6

66. Plutchik R. Emotion: A Psychoevolutionary Synthesis. NY Harper and Row. Harpercollins College Division; 1980.

67. Niolon R. List of Feeling Words [Internet]. [cited 9 May 2018]. Available: http://www.psychpage.com/learning/library/assess/feelings.html

68. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspect Psychol Sci. 2011; 6: 3–5. https://doi.org/10.1177/1745691610393980 PMID: 26162106

69. Hamilton WL, Clark K, Leskovec J, Jurafsky D. Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. 2016;

70. Fast E, Chen B, Bernstein M. Empath: Understanding Topic Signals in Large-Scale Text. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems—CHI '16. New York, New York, USA: ACM Press; 2016. pp. 4647–4657.

71. Mikolov T. Distributed Representations ofWords and Phrases and their Compositionality. NIPS Deep Learning Workshop. 2013. pp. 1–31.

72. Mikolov T, Yih W, Zweig G. Linguistic regularities in continuous space word representations. Proceedings of NAACL-HLT. 2013. pp. 746–751.

73. Kiss T, Strunk J. Unsupervised multilingual sentence boundary detection. Comput Linguist. 2006; 32: 485–525. https://doi.org/10.1162/coli.2006.32.4.485

74. Bird S, Bird S, Loper E. NLTK: The natural language toolkit NLTK: The Natural Language Toolkit. Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1. 2016. pp. 63–70.

75. Read J, Dridan R, Oepen S, Solberg LJ. Sentence Boundary Detection: A Long Solved Problem? Coling 2012. 2012. pp. 985–994.

76. Mnih A, Teh YW. A Fast and Simple Algorithm for Training Neural Probabilistic Language Models. Proceedings of the 29th International Conference on Machine Learning (ICML'12). 2012. pp. 1751–1758.

77. Řehůřek R, Sojka P. Software Framework for Topic Modelling with Large Corpora. Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks. 2010. pp. 46–50.

78. Thelwall M, Buckley K, Paltoglou G. Sentiment Strength Detection for the Social Web. J Am Soc Inf Sci Technol. 2012; 63: 163–173. https://doi.org/10.1002/asi.21662

79. Hutto CJ, Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. Eighth International AAAI Conference on Weblogs and . . .. San Francisco, CA, USA; 2014. pp. 216–225.

80. Bandaragoda TR, De Silva D, Alahakoon D. Automatic event detection in microblogs using incremental machine learning. J Assoc Inf Sci Technol. 2017; 68: 2394–2411. https://doi.org/10.1002/asi.23896

81. Donovan JL. Patient decision making. The missing ingredient in compliance research. Int J Technol Assess Health Care. 1995; 11: 443–55. https://doi.org/10.1017/S0266462300008667 PMID: 7591546

82. Stewart M. Towards a global definition of patient centred care. BMJ Br Med J. 2001; 322: 444–445.

83. Davis K, Schoenbaum SC, Audet A-M. A 2020 Vision of Patient-Centered Primary Care. J Gen Intern Med. 2005; 20: 953–957. https://doi.org/10.1111/j.1525-1497.2005.0178.x PMID: 16191145

84. Gerteis M. Through the Patient's Eyes: Understanding and Promoting Patient-Centered Care. In: eweb:124828. 1993.

85. Barry MJ, Edgman-Levitan S. Shared Decision Making—The Pinnacle of Patient-Centered Care. N Engl J Med. 2012; 366: 780–781. https://doi.org/10.1056/NEJMp1109283 PMID: 22375967

86. Elwyn G, Edwards A, Kinnersley P, Grol R. Shared decision making and the concept of equipoise: the competences of involving patients in healthcare choices. Br J Gen Pract. 2000; 50: 892–899. PMID: 11141876

87. Stacey D, Cl B, Mj B, Nf C, Kb E, Lyddiatt A, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. John Wiley & Sons, Ltd; 2011; 10. https://doi.org/10.1002/14651858.CD001431.pub3 PMID: 21975733

88. Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, et al. The impact of patient-centered care on outcomes. J Fam Pract. 2000; 49: 796–804. PMID: 11032203

89. White M, Dorman SM. Receiving social support online: Implications for health education. Health Educ Res. 2001; 16: 693–707. https://doi.org/10.1093/her/16.6.693 PMID: 11780708

90. Aiken LH, Clarke SP, Sloane DM. Hospital staffing, organization, and quality of care: Cross-national findings. Nurs Outlook. 2002; 50: 187–194. PMID: 12386653

91. Luxford K, Safran DG, Delbanco T. Promoting patient-centered care: a qualitative study of facilitators and barriers in healthcare organizations with a reputation for improving the patient experience. Int J Qual Heal Care. 2011; 23: 510–515. https://doi.org/10.1093/intqhc/mzr024 PMID: 21586433

Author/s:
De Silva, D; Ranasinghe, W; Bandaragoda, T; Adikari, A; Mills, N; Iddamalgoda, L; Alahakoon, D; Lawrentschuk, N; Persad, R; Osipov, E; Gray, R; Bolton, D

Title:
Machine learning to support social media empowered patients in cancer care and cancer treatment decisions

Date:
2018-10-18

Citation:
De Silva, D., Ranasinghe, W., Bandaragoda, T., Adikari, A., Mills, N., Iddamalgoda, L., Alahakoon, D., Lawrentschuk, N., Persad, R., Osipov, E., Gray, R. & Bolton, D. (2018). Machine learning to support social media empowered patients in cancer care and cancer treatment decisions. PLOS ONE, 13 (10), https://doi.org/10.1371/journal.pone.0205855.

Persistent Link:
http://hdl.handle.net/11343/271385

File Description:
Published version
License:
CC BY